



HELLENIC REPUBLIC
National and Kapodistrian
University of Athens
EST. 1837

Extending the YAGO4 Knowledge Graph with Geospatial Knowledge

Maria Despoina Siampou¹, Nikolaos Karalis², Manolis Koubarakis¹

¹ Artificial Intelligence Team, Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens

² DICE Group, Dept. of Computer Science, Paderborn University

5th International Workshop on Geospatial Linked Data at *ESWC 2022*

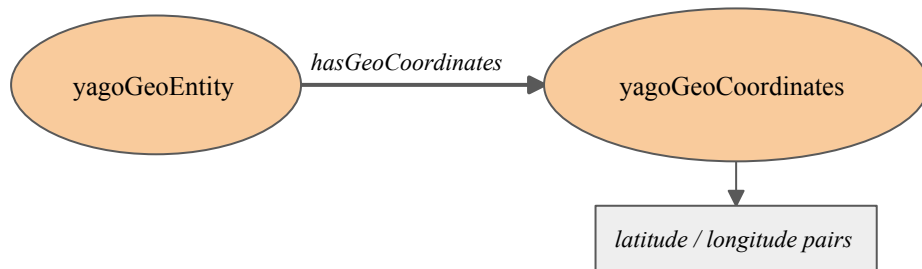
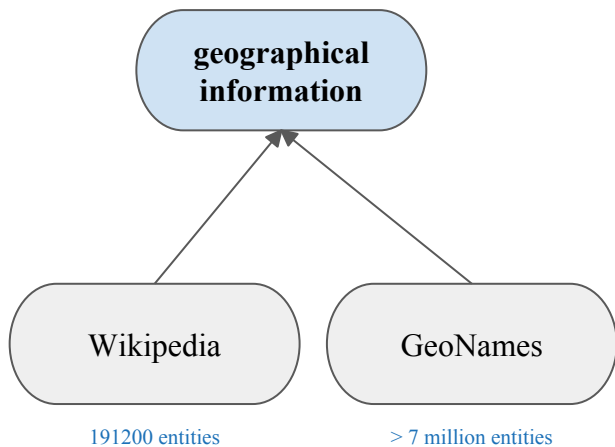
May 30, 2022

The evolution of YAGO



- **YAGO (2007)**
 - Wikipedia, WordNet
- **YAGO2 (2011)**
 - Spatial + Temporal Knowledge
 - WordNet, Wikipedia, GeoNames
 - SPOTL model
- **YAGO3 (2016)**
 - Multilinguality
 - non-English Wikipedias
- **YAGO4 (2020)**
 - schema.org, bioschemas, Wikidata

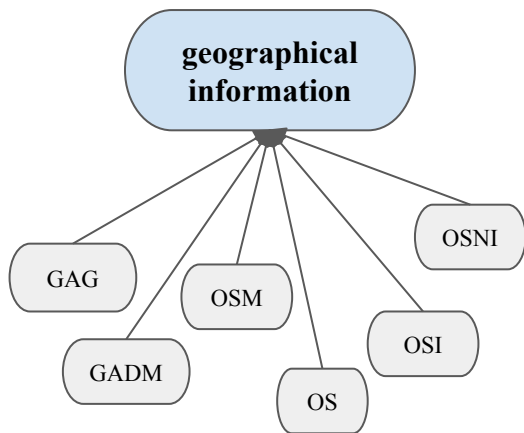
Geospatial Knowledge in YAGO2



```
@base <http://yago-knowledge.org/resource/> .  
<Athens> rdfs:label "Athens"@eng .  
<Athens> <isLocatedIn> <Greece> .  
<Athens> <hasLatitude> "37.979444444444745"^^<degrees> .  
<Athens> <hasLongitude> "23.71611111111113"^^<degrees> .
```

“knows only about coordinates”

Geospatial Knowledge in YAGO2geo



“introduces 137K linestrings and 640K polygons and multipolygons”

1. OGC Geometries

- eg.: polygons, multipolygons, linestrings

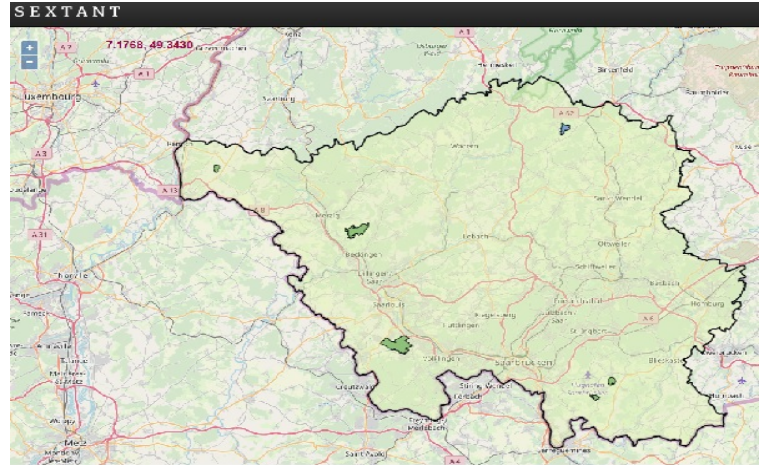
2. Topological relations, OGC Vocabulary

- eg.: sfTouches, sfWithin

```
@base <http://yago-knowledge.org/resource/> .
<Athens> rdfs:label "Athens"@eng .
<Athens> <isLocatedIn> <Greece> .
<Athens> <hasLatitude> "37.979444444444745"^^<degrees> .
<Athens> <hasLongitude> "23.71611111111113"^^<degrees> .
<Athens> geo:sfTouches <Kaisariani> .
<Athens> geo:hasGeometry y2geo:Geometry_gag_9186.
y2geo:Geometry_gag_9186 geo:asWKT "MULTIPOLYGON((( ... )))"
```

Querying YAGO2geo

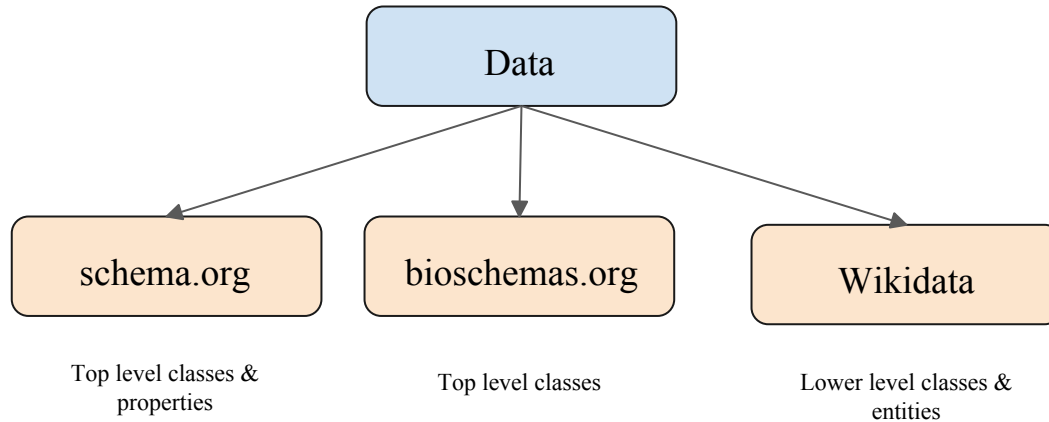
“Find the geometries of water bodies and forests that are within Saarland.”



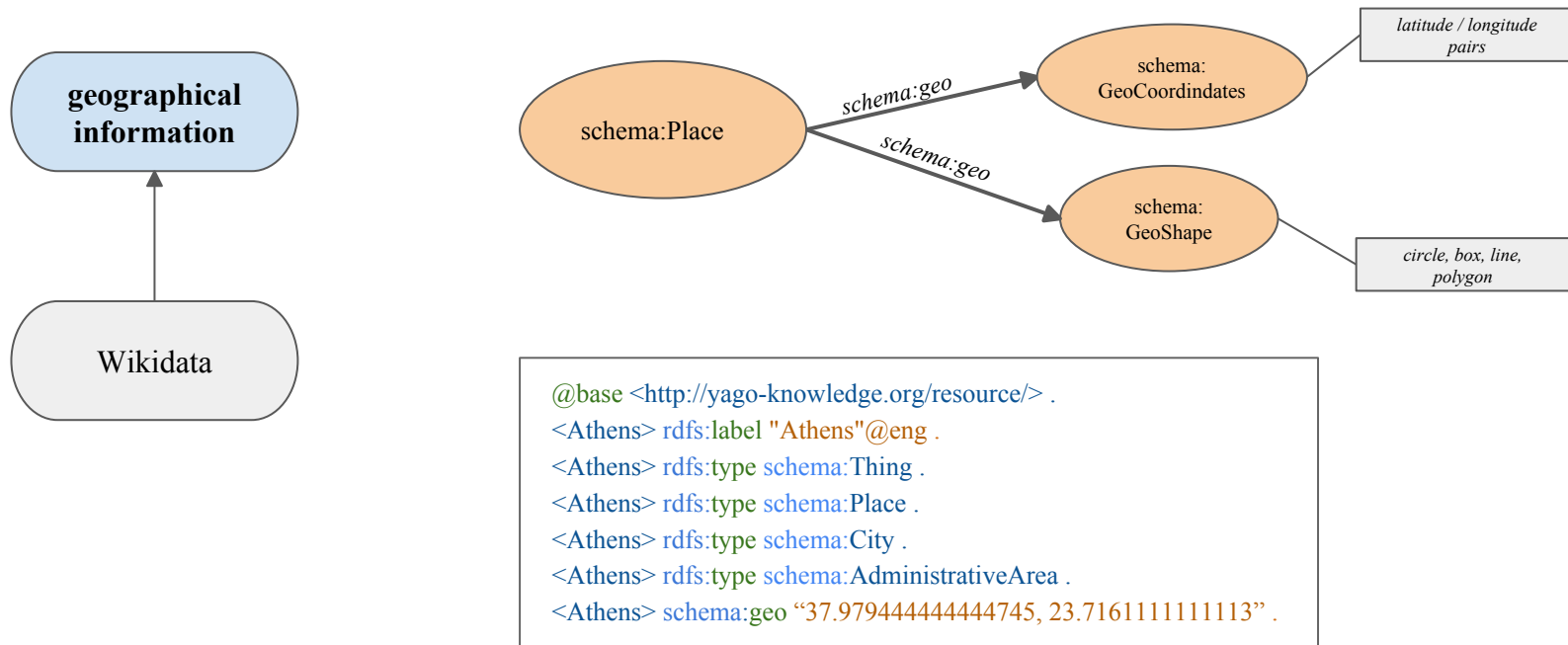
¹ Strabon: <http://test.strabon.di.uoa.gr/yago2geo/>

² Sextant: <http://sextant.di.uoa.gr/>

The Knowledge Graph YAGO4



Geospatial Knowledge in YAGO4



“knows only about coordinates”

Motivation

- Not enough geospatial knowledge in current KGs
- Importance of geospatial data in today's tasks & applications
 - data driven decisions, spatial analysis
- Spatial representation in YAGO4
 - location representation using geometry centers

“Which is the largest forest in Greece?”

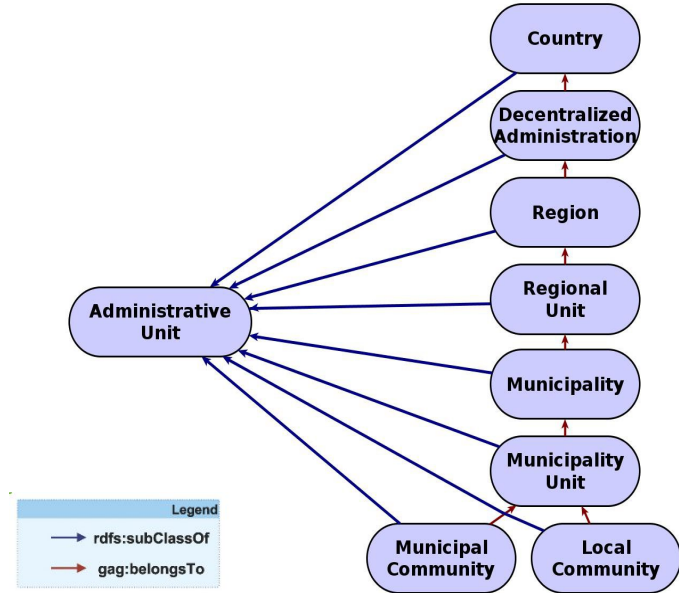
“Which are the neighboring municipalities of the municipality of Athens?”

“Which municipalities belong to the administrative area of Attica?”

Question

“How can we extend YAGO4 to support richer and more precise geospatial knowledge?”

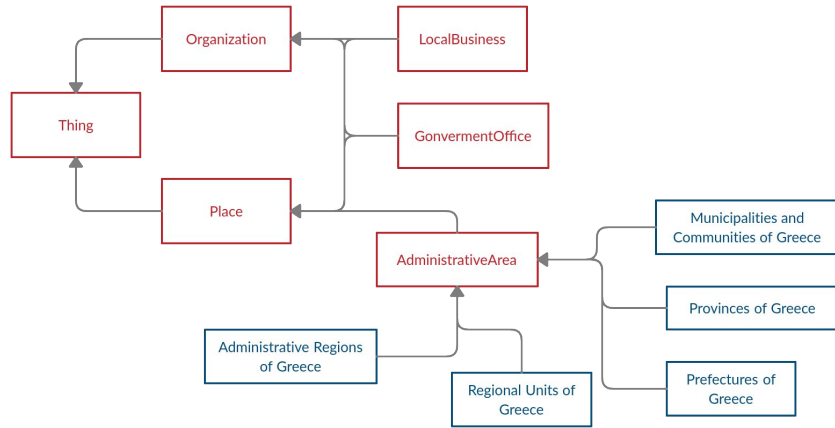
Data Sources; *Greek Administrative “Geography” Dataset (GAG), Kallikratis Law (2010)*



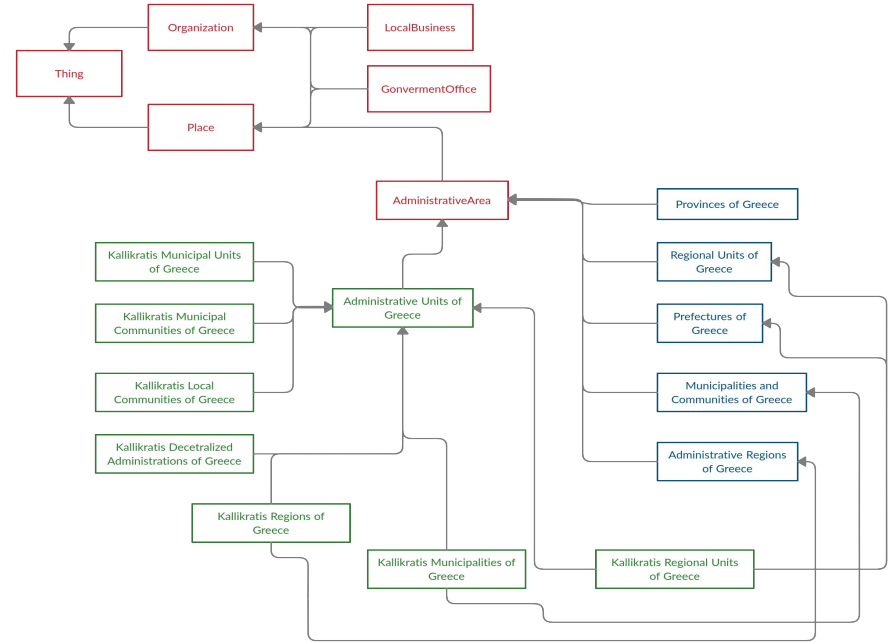
Data Sources; *Comparison of instances*

Administrative Unit	# instances in GAG	# instances in YAGO4
Decentralized Administrations	7	7
Regions	13	13
Regional Units	74	77
Prefectures	0	55
Provinces	0	70
Municipalities	325	340
Municipal Units	1034	0
Municipal Communities	3	0

Methodology; *Taxonomy Extension*



*Administrative organization of Greece
in YAGO4 taxonomy*



*Administrative organization of Greece
in YAGO4 extended taxonomy*

Methodology; *Properties Addition*

→ Object Properties

- ◆ has_seat

→ Data Properties

- ◆ has_population
- ◆ has_code

```
@base <http://yago-knowledge.org/resource/> .  
<Athens> rdfs:label "Athens"@eng .  
<Athens> rdfs:type schema:Thing .  
<Athens> rdfs:type schema:Place .  
<Athens> rdfs:type schema:City .  
<Athens> rdfs:type schema:AdministrativeArea .  
<Athens> y4geo:has_population "3218218"^^xsd:integer .  
<Athens> y4geo:has_code "8133876"^^xsd:integer .
```

Methodology; *Matching Phase*

1. Label Similarity Filter

- Matches entities based on their name literal
- Metric: *Levenshtein distance*
- Threshold t_a : 0.8

2. Geometry Distance Filter

- Matches entities based on the distance between the center of the candidate geometry and the coordinates of the entity
- Metric: *Euclidean distance*
- Threshold t_b : 25 km

! Avoid duplicate information

References: *YAGO2*, *LinkedGeoData*, *YAGO2geo*

Methodology; *Matching Phase*

1. Matched Entities

- Extend entity with new information

2. Unmatched Entities

- Introduce new entity into YAGO4

! Avoid duplicate information

```
@base <http://yago-knowledge.org/resource/> .
<Athens> rdfs:label "Athens"@eng .
<Athens> rdfs:type schema:Thing .
<Athens> rdfs:type schema:Place .
<Athens> rdfs:type schema:City .
<Athens> rdfs:type schema:AdministrativeArea .
<Athens> y4geo:has_population "3218218"^^xsd:integer .
<Athens> y4geo:has_code "8133876"^^xsd:integer .
```

matched entity

```
@base <http://yago-knowledge.org/resource/> .
<kallikratis_entity_304> schema:alternateName "ΔΗΜΟΤΙΚΗ ΕΝΟΤΗΤΑ ΜΕΣΣΑΠΙΩΝ" .
<kallikratis_entity_304> rdfs:type schema:Thing .
<kallikratis_entity_304> rdfs:type schema:Place .
<kallikratis_entity_304> rdfs:type schema:AdministrativeArea .
<kallikratis_entity_304> y4geo:has_population "13756"^^xsd:integer .
<kallikratis_entity_304> y4geo:has_code "914802"^^xsd:integer .
```

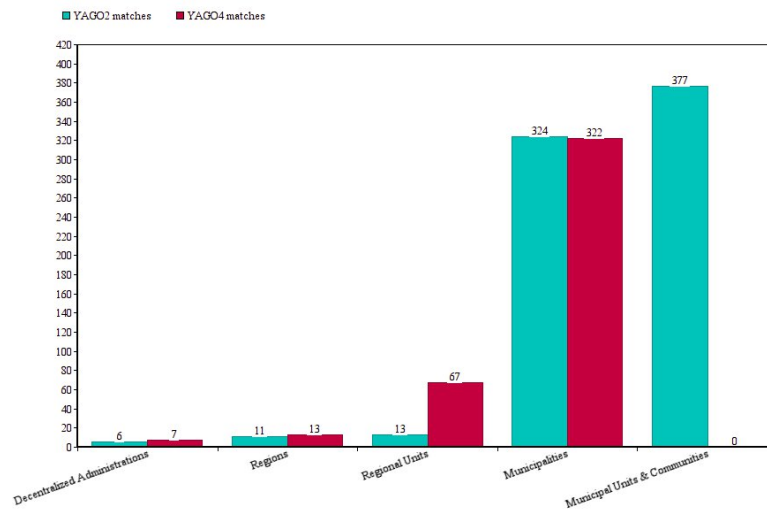
unmatched entity

Results; *Matched instances*

GAG Classes	YAGO4 Classes	Number of matched GAG entities
Decentralized Administrations	Government Offices	7/7
Regions	Regions of Greece	13/13
Regional Units	Regional Units and Prefectures of Greece	67/74
Municipalities	Municipalities and Communities of Greece	322/325

Results; *Comparison with YAGO2geo*

- Different class hierarchies
 - different number of matches
- YAGO2geo
 - absent entities in YAGO2
 - inability to match the majority of regional units
- YAGO4geo
 - taxonomy is more compatible to our task
 - absent entities in YAGO4
 - due to absence in Wikidata
 - inability to match municipal units & communities
 - duplicate entities in YAGO4
 - duplicate entities representing regional units & municipalities



Results; *Location Coordinates*

```
POLYGON ((  
21.71163060727338 40.85096274759539,  
21.71163660172085 40.85091782978087,  
... ,  
21.71163060727338 40.85096274759539 ))
```

WKT polygon

```
"  
21.71163060727338, 40.85096274759539  
21.71163660172085, 40.85091782978087  
...  
21.71163060727338, 40.85096274759539  
"
```

schema polygon

Limitations

- Schema.org does not provide support for complex geometries (*eg. multipolygons*)
- Schema.org has its own geometry serialization



Partial Solutions

Results; *Location Coordinates*

Partial Solutions

1. Deconstruct complex geometries
 - Split multipolygons to individual polygons
 - Introduce *excludeGeoShape* property and use aggregation lists
 - **Requires extra effort from the webmasters**

1. Utilize *schema:additionalProperty* property
 - Can represent additional characteristics
 - **New information cannot be properly exploited**



Inability to perform proper GeoSPARQL queries

```
@base <http://yago-knowledge.org/resource/> .
<Athens> rdfs:label "Athens"@eng .
<Athens> rdfs:type schema:Place .
<Athens> schema:geo "21.71163060727338, 40.85096274759539 ...
                    21.71163060727338, 40.85096274759539" .
...
<Athens> schema:geo "23.12134677656756, 56.73827386778883 ...
                    23.12134677656756, 56.73827386778883" .
```

```
@base <http://yago-knowledge.org/resource/> .
<Athens> rdfs:label "Athens"@eng .
<Athens> rdfs:type schema:Place .
<Athens> schema:additionalProperty y4geo:kallikratis_geometry_18 .

y4geo:kallikratis_geometry_18 rdfs:type schema:propertyValue .
y4geo:kallikratis_geometry_18 schema:name "Kallikratis_Geometry_18" .
y4geo:kallikratis_geometry_18 schema:value
"MULTIPOLYGON(((...)))" .
```

Conclusions

- schema.org provides a simple and easily maintained taxonomy **but** :
 - provides a less powerful feature/geometry representation
 - does not cater for :
 - well-known spatial literal formats (*eg. KML, GeoJSON, WKT*)
 - complex geometries (*eg. multipolygon, geometry collection*)
 - multiple coordinate reference systems
- YAGO4 inherits these limitations due to its class hierarchy alignment to schema.org.

Open Issues & Recommendations on schema.org repository

- Addition of *GeospatialGeometry* property to schema.org
 - *supertype of GeoShape*
 - *accommodate definitions from geoSpatial practices*
 - *models spatial relations from DE-9IM (e.g. geoEquals, geoCrosses, geoOverlaps etc.)*

- Align geometry serialization to WKT and GeoJSON
 - *e.g. commas within pairs of coordinates instead of spaces*
 - *e.g. introducing WKT as a subtype of Text*

Thank you!